

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: April 26, 2003, 13:02:26 ; Search time 79 Seconds  
(without alignments)  
1405.035 Million cell updates/sec

Title: US-10-027-000-2  
Perfect score: 4391  
Sequence: 1 MADIVEALIKLTLAEKVD.....DGVLRGRFTVGETYMGV 833

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	969	22.1	769	21 AAB18648	Amino acid sequenc.
2	969	22.1	769	21 AAB18648	ORF 11 encoded par
3	953	21.7	721	19 AAM4862	Thermotoga maritima
4	951.5	21.7	3782	21 AAY71179	S. venezuelae deso
5	944.5	21.5	808	21 AAE24228	Streptomyces venez
6	944.5	21.5	809	21 AAY77189	S. venezuelae macr
7	907	20.7	721	18 AAE24237	Streptomyces venez
8	869	19.8	752	17 AAM34558	Thermotoga maritima
9	869	19.8	752	17 AAM34558	Chimeric thermos
10	860	19.6	735	19 AAM69761	Acetobacter xylinu

11	796.5	18.1	803	16 AAB85197	Tomatinase - a sap
12	747	17.0	778	19 AAM35004	Thermotoga maritima
13	731.5	16.7	744	13 AAR25384	Trichoderma reesei
14	731.5	16.7	744	21 AAB08340	Amino acid sequenc
15	731.5	16.7	744	21 AAY36548	Trichoderma reesei
16	728.5	16.6	793	16 AAB85199	Avenacinase-like p
17	698	15.9	793	16 AAB85200	Avenacinase-like p
18	681	15.5	792	16 AAB85198	Avenacinase - a sa
19	620	14.1	801	21 AAY77205	S. venezuelae deso
20	603.5	13.7	841	22 AAB82320	Aspergillus niger
21	603.5	13.7	860	22 AAB82320	Aspergillus niger
22	603.5	13.7	880	22 AAB82327	Aspergillus niger
23	603.5	13.7	883	22 AAB82328	Aspergillus niger
24	572.5	13.0	858	23 AAB07868	Herbicidally activ
25	572.5	13.0	858	23 AAB90815	Herbicidally activ
26	564	12.8	756	23 ABA46400	Herbicidally activ
27	541.5	12.3	723	23 ABA46400	Herbicidally activ
28	532.5	12.1	774	23 ABB93754	Herbicidally activ
29	518.5	11.8	781	23 ABB92368	Herbicidally activ
30	512.5	11.7	868	21 AAG36535	Herbicidally activ
31	495.5	11.3	773	23 ABB93409	Herbicidally activ
32	495.5	11.3	1476	21 AAG49341	Herbicidally activ
33	491.5	11.2	784	23 ABB93421	Herbicidally activ
34	485	11.0	784	23 ABB93421	Herbicidally activ
35	457.5	10.4	739	21 ABB91643	Herbicidally activ
36	456.5	10.4	767	23 ABB91643	Herbicidally activ
37	455	10.4	728	21 AAG36537	Herbicidally activ
38	433.5	9.9	804	18 AAM10033	Human polypeptide
39	426	9.7	1328	21 AAG49342	Novel human diagno
40	417.5	9.5	342	23 ABB93023	Aspergillus oryzae
41	369.5	8.4	233	22 ABB93023	Herbicidally activ
42	350.5	8.0	798	21 AAT52699	Herbicidally activ
43	333	7.6	650	23 ABB92769	Herbicidally activ
44	322.5	7.3	608	21 AAG54268	Herbicidally activ
45	322.5	7.3	608	23 ABB92570	Herbicidally activ

## ALIGNMENTS

RESULT 1  
AAB18648  
ID AAB18648 standard; Protein: 769 AA.

AC AAB18648;

DT 22-JAN-2001 (first entry)

XX Amino acid sequence of an ORF11 beta-glucosidase (dear).

KW Narbonolide synthase; polyketide synthase gene; narbonolide synthase;  
KW antibiotic; C12-hydroxylase; pick; desosamine biosynthesis;  
KW desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;  
KW plicomycin biosynthesis; beta-glucosidase.

XX Streptomyces venezuelae.

OS US6117659-A.

PN 12-SEP-2000.

PD 27-MAY-1999;

PF 99US-0320878.

PR 28-MAY-1998;

PR 22-SEP-1998;

PR 08-FEB-1999;

PR 20-MAY-1999;

PR 30-APR-1997;

PR 06-MAY-1998;

PR 28-AUG-1998;

(KOSA-) KOSAN BIOSCIENCES INC.

PI Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;

XX WPI: 2000-610844/58.

PT New recombinant pick hydroxylase gene of *Streptomyces venezuelae* useful  
PT for converting ketolides to antibiotics and as antibiotics and  
PT intermediates in the synthesis of compounds with pharmaceutical value

XX Disclosure: Columns 37-38; 117pp; English.

XX The present sequence represents a beta-glucosidase polypeptide.  
CC The nucleotide sequence encoding it is used in the course of the  
CC invention. The specification describes a recombinant DNA compound  
CC expressing recombinant polyketide synthase genes in host cells for the  
CC production of narbonolide, narbonolide derivatives and polyketides that  
CC are useful as antibiotics, and as intermediates in the synthesis of  
CC compounds with pharmaceutical value. The DNA compounds may also encode  
CC a 412-hydroxylase (pick), desosamine biosynthesis and desosaminyl  
CC transferase enzymes (useful for conversion of ketolides to antibiotics),  
CC and the beta-glucosidase enzyme (involved in picromycin biosynthesis).  
CC These compounds are also useful for increasing the antibiotic activity  
CC of a compound relative to the unhydroxylated compound. The recombinant  
CC host cells are useful as genetic systems that allow rapid engineering  
CC of the narbonolide polyketide synthase. These would be valuable for  
CC creating novel ketolide analogs for pharmaceutical applications.

XX Sequence 769 AA;

Query Match 22.1%; Score 969; DB 21; Length 769;  
Best Local Similarity 32.6%; Pred. No. 1,5e-77;

Matches 277; Conservative 112; Mismatches 340; Indels 120; Gaps 25;

13 LTLAEKVDLAGIDFW-----HTKALPHGVPSLFTDGPNGVGTFFGVPA 62  
1 MTLDEKISFV-----HMLDLPDRONVGYLPGVPRIGIPLRADGPNGIR----LVGQAT 52  
63 CFCPGTSLGTFNQTLLLEAGKMKREKALAKSAHVILGPTIMKQSPILGRFESTIGDP 122  
53 ALPAPVALASTFDQTMADSGYKMGDRDALNDWVLGPMNNIRPHGRNYETFFSBDP 112  
123 FLAGLGAALIRIGTQVQATIKHFLCNDQEDRRMVOISITERALREIYALPQIIVR 182  
113 LVSSRTAVAOIKIGAGMTTAKHFAANNOENRFSVANVDEQTLREIEPRAE-ASS 171  
183 DSQGAFTAYNGINGVSCSENPXYDGLRKEMGMDGLMSDWGTGSTEAVVAGIDL 242  
172 KAGAGSMCAVYNGLNKPCSGNDELINVLKQMGFGQWMSDWLDT-PGDDATIKTKIDQ 230  
243 EM-----PGPP-RFRGETLKFVNSNGK-PFIHVIDQARAREVLOFVKCAASGV 288  
231 EMGVLELPDVKPEPSPAKFFGEALKTAVLVNGTVPEAAVRSARITVQMEKEGGLLAT 290  
289 TENGPEPTVNNPETAALLRKVNGNGIYILKNNVLLS--KKKTLIVSPNAKQATYH 346  
291 PAPRPE---RDKAGAOVASRKVAENGAVLLRREGQALPLADAGCSIAVITPAVDPKVT 347  
347 GGGSAAALRAYVATPFDGLSKOLETPSYTGAATVTPVIGEGOLPPDGAQPKRWAVEN 406  
348 GIGSAHVVDASAAPLDITIKAR-----AGAGATVYIEGDETFGTQIPAGWLSAPFN 399  
407 EPPGPTRNRHIDELFTTDMHLVDYIPKKAADTWYADMEGTYTADEDCYELGLVVCGT 466  
400 QC-----HOLE---PGKAGALY--DQTLVVPADGEVRIAVRATG- 433  
467 AKAYVDOLVVDNATKQVPGDAFPGSATREBTGRINLVKNGTYFKTEFSPAPRTYTLKCD 526  
434 --GYATVQL---GSHTEAQQVYKYS--SPLKLTGK--THL-----TISGF 473  
527 TIVPHGSLRRGGCKVVIDQAEIKSVYALAKENHDQVITICAGLNADWETEGADBSMKLIP 586  
474 AMSATPLSLIEIGWVTPAADAATIAKAVESARKARTAVVFA---YDQTEBVDPRNLSLR 530

QY 587 VLDQLIADVAANPNTVVYVMTGPEENRPMUDATPAVIOAWNGENGTNSIADVFCDYN 646  
DB 531 TODKLISAADANPNTIVVLTNGSSVLMKPMVSKTRAVLDMMYPPQAAAEKFTALLIYDVN 550  
QY 647 PSGKLSLSPKRLQONPAFLNFRTEAG-----RTLYGEDVYVGYRYEPADKDVN 656  
DB 591 PSGKLTQSF-----PAENQHAVAGDPITSYPCVDNQOTYREGIHGVHMFKEKENVPL 643  
QY 697 FPFHGLSYTTFEAFSNLSVSH-KDGKLSVLSKNGTSVPGAOVYKPLDAAKINR 755  
DB 644 FPFHGLSYTSPQSAFVAVTSIGLKYTVYVANSKRAGQEVYVAYLQASRNTVAPQA 703  
QY 756 VKELGKPAKVELQCEFTKAVTIEQEKYVAAYFDEERDQCVKRGDYEVYVSDSSAKDG 815  
DB 704 KKKLVGYKVSILAAGEAKTVVNVDRQL-QFMDAATDMKKTGTGRNLLQOTGSSA----- 758  
QY 816 VALRKGFTV 824  
DB 759 -DLRGSATV 766

# RESULT 2

AAV67212 standard; protein; 769 AA.

AAV67212;

23-MAR-2000 (first entry)

ORF 11 encoded partial beta-glucosidase, deasr amino acid sequence.

Narbonolide polyketide synthase: PKS; desosamine biosynthesis; deasr;  
beta-glucosidase; antibiotic production; narbomycin; picromycin;  
ketolide.

*Streptomyces venezuelae*.

WC9961599-A2.

02-DEC-1999.

27-MAY-1999; 99WO-US11814.

28-MAY-1998; 98US-0087080.

28-AUG-1998; 98US-0141908.

22-SEP-1998; 98US-0100880.

08-FEB-1999; 99US-0119139.

(KOSA-) KOSAN BIOSCIENCES INC.

Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;

WPI: 2000-072618/06.

N-PSDB; AA256002.

New recombinant DNA encoding a domain of narbonolide polyketide  
synthase, for production of ketolide antibiotics -

Example 2; Page 33; 98pp; English.

This is the *Streptomyces venezuelae* partial beta-glucosidase, deasr amino  
acid sequence, encoded by ORF11. This protein is involved in desosamine  
biosynthesis. The invention relates to recombinant DNA containing a  
coding sequence for a narbonolide polyketide synthase (PKS). Polyketides  
are compounds synthesized from 2-carbon units through a series of  
condensations and subsequent modifications. Modular PKSs are responsible  
for the production of many antibiotics including picromycin. The  
narbonolide PKS consists of a loading module, six extender modules, and  
two thioester domains. Four proteins make up the narbonolide PKS (PICAI,  
PICAIL, PICAIIT and PICAIIV). PICAI includes the loading module and  
extender modules 1 and 2, PICAIIT includes extender modules 3 and 4,  
PICAIIV includes extender module 5 and PICAIIV includes extender module 6  
and a type II thioesterase domain. The second type II thioesterase

CC domain is found on the PCB protein. The nucleotide sequences encoding  
 CC all of these proteins can be isolated in recombinant form from the  
 CC recombinant cosmid PK05023-27 (see AA256001). Naponolide is  
 CC desosaminylated in *S. venezuelae* to yield narbomycin, the desosaminyl  
 CC transferase enzyme is required for this conversion, and the desosamine  
 CC biosynthetic genes are also found in cosmid PK05023-27. The recombinant  
 CC DNA of the invention is used to express, in transformed cells,  
 CC naponolide (or its derivatives) or other ketolides (particularly  
 CC hybrids), which may then be converted (e.g. by other enzymes  
 CC recombinantly expressed in the same hosts) to polyketide antibiotics or  
 CC their intermediates. The antibiotics are useful in human or veterinary  
 CC medicine.

CC Sequence 769 AA:

Query Match 22.1%; Score 969; DB 21; Length 769;

Best Local Similarity 32.6%; Pred. No. 1.5e-77;

Matches 277; Conservative 112; Mismatches 340; Indels 120; Gaps 25;

13 LTLEKVDLLAGIDFW-----HTKALPKHGVPSLRFDGPNVGRTKFGNVPAA 62  
 Db 1 MTLDEKISV-----HWALDPDQONVGLPGVRLGIPELRADGPNGR-----LVGQRT 52  
 QY 63 CPECGTSLGSTEQTLLLEBAGKMKKEALAKSAHVILGPTIMORSPLUGRGFESTIGEDP 122  
 Db 53 ALPAPALASTFDDTMDASYKMGGRDALNODMVLGPMNNIRVPHGGRNYETFESEDP 112  
 QY 123 FLAGLGAALIRIGTGTGATIKHFLCNDQEDRRMVSITYERLKRITVLPQIAR 182  
 Db 113 LVSSRTAVVQIKIGDAGLTAKHFAANNOENRFSVANVDEQTLREIEPAEE-SS 171  
 QY 183 DSQGFAMTAAYNGINVCSENPXYLDGMLRKEMGDLGMSDMWGTYSSTEAAGLDL 242  
 Db 172 KAAGSFMCAIYNLNKSPSCGNDELLNNVLRQMGFGQVMSDMLAT--GCTAIIRKGLDQ 230  
 QY 243 EM-----PGPP--RFRGETLKFVNSNGK--PFIHVLDQRAKEVLOFVKCAASV 288  
 Db 231 EMGVELPGDVPKGESPAPKPFGEALKTAVLNGTPEAATVSAERIVGOMKEFGILLAT 290  
 QY 289 TENGPEPTVANNPETAALLRKVNEGIVLKNENNVLPSS--KKKTLVGNNAQATIH 346  
 Db 291 PADRPE--RDKAGAAVSRKVAENGAVLLRNEGALPLAGAGSIAVIGPTAADPKYT 347  
 QY 347 GGGSALRAYAVVTPPDGSKOLETPPSYTVGAYTVPPIEGECLTPDGAAGMRMVEN 406  
 Db 348 GLGSAHVVDASAAPLDITIKAR-----AGAGATVTTETGEETFGTIPAGNLSPAFN 399  
 QY 407 EPPGTENRQHIDELFTKTDMLVDYHPRKADTWYADMEGTYTADDECTYELGLVVCST 466  
 Db 400 QG-----HGLE---PGKAGALY---DGLTVPADGEYRIAVRATG- 433  
 QY 467 AKRYVDOLVVDNATKOVGDAPFGSATBEGRINLVVGNTRYKFIETGSAPTTLKGD 526  
 Db 434 -GYATVOL-----GSHITBAGOVYKVS--SPLKLTFTG--NHKL-----TISGF 473  
 QY 527 TIYPGHGSLRVGGCKVIDDOAEIEKSVLAKEDHWIITAGINADMETEGDARASKLPG 586  
 Db 474 AMSATFLSLELGWTPPAADATITAKAVESAKRARTAVNVA---YDDGTGCVDRPNLSLPG 530  
 QY 587 VLDOLIAVAAANPNTVVVWQGTPEEMFWLDATPAVIGAWYGGNETGNSIADVYFGDN 646  
 Db 531 TQCKLISAVADANPNTIVLVNTGSSVLMWLSKTRAVLDMWYPGQAGABATVLAALYGDVN 590  
 QY 647 PSGLKLSFPKRIQDNPAFLNRTENG-----RLVGEDVYVVGRIYEFADKDVN 696  
 Db 591 PSGLKLSFPKRIQDNPAFLNRTENG-----RLVGEDVYVVGRIYEFADKDVN 643  
 QY 697 FPPGHGSLSTTFEASNLISVSH--KDGLSYSLSVKNTGSPGCAOVOLYKPKQAAKINRP 755  
 Db 644 FPPGHGSLSTTFEASNLISVSH--KDGLSYSLSVKNTGSPGCAOVOLYKPKQAAKINRP 703  
 QY 756 VKELKFAKVELOPGETKAVTTEDEKRYAAVFEDEEDQWCVKGDYEVIVSDSSAAKDG 815  
 Db 756 VKELKFAKVELOPGETKAVTTEDEKRYAAVFEDEEDQWCVKGDYEVIVSDSSAAKDG 815

Db 704 KKILVYTKVSLAAGEAKTVTVNDRRL-QFWDAIDDMKTKGTGNLLQTGSSA----- 758  
 QY 816 VALRGEKTV 824  
 Db 759 -DLKRSATV 766

RESULT 3  
 ID AAM49862  
 AC AAM49862  
 DE AAM49862; standard; Protein; 721 AA.

XX 21-DEC-1998 (first entry)

XX Thermotoga maritima MSB8-6G glycosidase.

KW Glycosidase; MSB8-6G; thermostable enzyme; oligosaccharide;  
 RV glucose; sugar; baking; textile; detergent; beta-galactosidase.

XX Thermotoga maritima strain MSB8-6G.

OS W09824799-A1.

PN 11-JUN-1998.

PD 08-DEC-1997; 97WO-US22623.

PR 10-OCT-1997; 97US-0949026.

PR 06-DEC-1996; 96US-0056916.

PA \* (DIVE-) DIVERSA CORP.

PI Bylina EJ, Lam DE, Mathur EJ, Swanson RV;

DR WPI: 1998-362407/31.

DR N-PSDB; AAV36911.

PT Glycosidase enzymes from organisms of the genera *Staphylothermus*,  
 PT *Pyrococcus* and *Thermococcus* - for deriving sugar from  
 PT oligosaccharides, useful in the e.g. food processing, textile or  
 PT baking industries

PS Claim 1; Fig 5a-b; 92pp; English.

CC This is the amino acid sequence of glycosidase MSB8-6G, deduced  
 CC from a polynucleotide (see AAV36911) of a *Thermotoga maritima* MSB8  
 CC clone (66) that grows optimally at 85 degC in high salt medium.  
 CC The sequence shows 45% amino acid identity to beta-galactosidase  
 CC bglB of *Clostridium thermocellum*. The invention provides 18  
 CC polynucleotides (see AAV36907-24) coding for thermostable glycosidases  
 CC (see AAM49858-75) having glucosidase, alpha-galactosidase,  
 CC beta-galactosidase, beta-mannosidase, beta-mannanase, endoglucanase  
 CC or pullulanase activity. Vectors and host cells are also claimed.  
 CC A method is provided for producing the enzymes by recombinant  
 CC techniques. A claimed method for generating glucose from soluble  
 CC cell oligosaccharides comprises contacting a sample (selected from  
 CC dairy products, fruit juice, detergent, textile, guar gum, animal  
 CC feed, plant biomass or waste product) containing oligosaccharides  
 CC (selected from maltose, cellobiose, lactose, sucrose, raffinose,  
 CC stachyose, verbascose, cellulose, starch, amylose, glycogen,  
 CC disaccharides, polysaccharides and pullulan) with one of the  
 CC claimed glycosidases such that glucose is produced.

XX Sequence 721 AA:

Query Match 21.7%; Score 953; DB 19; Length 721;

Best Local Similarity 30.3%; Pred. No. 3.7e-76;

Matches 261; Conservative 127; Mismatches 286; Indels 186; Gaps 25;

QY 6 VEAILKKTITLAEKVDLLAGIDF-----WHTKALPKHGVPSLRFDGPNVG 50  
 Db 4 IDEILSOLTEEKVLLVVGILPGILGNPHSRVAGAGETHPVRLGIAPVAVLADGAGL 63

QY 51 R-----GTFKFNQVPAACPCGTSLGTFNQTLLEBAGKMGKEALAKSAHVILGPTINMO 106  
 DB 64 RINPTRENDENTYTTTAFVEIMLSTNRDLLEEVGAMGEVEYEGVDILLAPAMNIIH 123  
 QY 107 RSPILGREGESIGDEPFLAGLAALINGIQTSTQVATIKHFLCDOEDRRMNOVSITYE 166  
 DB 124 RNPICGRNFEYSSEDPVLSGEMASAFVKGVSQGVGACIKHENVANNOENRNVVDITYSE 183  
 QY 167 RALREIYALPQIVARSDSPGAFMTAYNGINGVSCSENPBYDGLRREMGDGLMSDW 226  
 DB 184 RALREIYALKGEIYAKKAPPTVMSAYNKLKNCYCSQNMELKKYLREEMGFGGVMSDW 243  
 QY 227 YGTYSTTEAVVAGLLEMPGP-----RFRGETLKFNVNSGKPFIVHIDQAREVLQF 279  
 DB 244 YAGDNPVEQLKAGNMIMGKAYQVNTERRDEIEIMELKRGKLSSEEDLDECVANILK- 302  
 QY 280 VKKCAASVTENGPE---TTVNNTPET---AALLKRVNEGIVLLKNNENYPLSKKKKT 333  
 DB 303 -----VLNAPSPFKGYRNSKPDLSHAENVAYEAGEGVLLDEN-NGVLPFDENTHV 353  
 QY 334 LTVGPNAKQATYHGGGSAALRAYVAVTPDGLSKQLETPSTVGVATYTPPILEGQCLT 393  
 DB 354 AVFGIGQITITGCTGSGGTHPRYITISLEGIKE----- 387  
 QY 394 PDGAPMRWRVNEPPGPTRNQHIDELFTKIDMLVDYHPRKADTWADMEGYTDAE 453  
 DB 388 -----RMKFDDELASTYEYIKK-----MRETEYKPR-TDSW----- 420  
 QY 454 DCTYELGLVCGTAKAYVDQLYVNNATKQVPGDAFPGSATPEETGRINLYKGNTRYKFI 513  
 DB 421 -----GTVI---RKPLPENFLSEKRIKPKKNDVAVV-----ISRLSGEYDKR- 463  
 QY 514 EFGSAPTYLKGDJTYIPGHSILVGGCKYIDDOAEIEKSVALAKE-HQVITICAGLNADW 572  
 DB 464 -----PVKGDYFIS-----DDELILIKTV-SKEFHQ----- 489  
 QY 573 ETBEGADRASKMLPGVLDOLIAVNAANPNTVVMOTGPPEEM-PVLDATPAVIOAWYGN 631  
 DB 490 -----GKKVVVLLNIGSPIEIVASWRDLDVGDILVWQAGQ 523  
 QY 632 EGNISADVPGDYNPSGKLSLSPFKRLQDNEAFL---NFRTEARLTLYGDEVYGYRY 688  
 DB 524 EMGRITADVLRKINSGLPTTFPRDSDVSWTFPGEPKPNOPRVYIEEDYIGIRY 583  
 QY 689 EFADKDVNPPFGHLSYTFEAFSNLSVSHKDKLSVSKVTKGVSVPAGVQAOLYVPLQ 748  
 DB 584 DIFGVPAPEYFGYGLSYTFEYKDKLIALDGFTLAVSYITNTGDRACKESQVYIK-AP 642  
 QY 749 AAKINRPVELKGFAPKE-LDPGETKAVTIEQEKYVVAAYFDEERDQCKVEKGYEVIVS 807  
 DB 643 KGRINKPQELKAFKHTKLLNPGSESEISLEIPLRLASFDGKE---WVESGEYEVYRG 699  
 QY 808 DSSAADGVALRGKFTV-GE 826  
 DB 700 ASSR---DIRLDDIFLVEGE 716  
 RESULT 4  
 AAY77179 standard: Protein; 3782 AA.  
 ID AAY77179 standard: Protein; 3782 AA.  
 AC AAY77179;  
 XX  
 DT 05-JUN-2000 (first entry)  
 DE S. venezuelae desosamine biosynthetic plkB-encoded protein, SEQ ID NO.4.  
 XX  
 KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;  
 KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;  
 KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;  
 KW chronic obstructive pulmonary disease; respiratory inflammation;  
 KW hypercholesterolemia; crop protection agent.

XX OS Streptomycetes venezuelae ATCC15439.  
 XX  
 XX W0200000620-A2.  
 XX  
 XX 06-JAN-2000.  
 XX  
 XX 25-JUN-1999; 99WO-US14398.  
 XX  
 XX 26-JUN-1998; 98US-0105537.  
 XX  
 XX (MINU) UNIV MINNESOTA.  
 XX  
 XX Sherman DH, Liu H, Xue Y, Zhao L;  
 XX  
 XX WPI: 2000-160679/14.  
 XX  
 XX DR N-PSDB: AA287284.  
 XX  
 XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.  
 XX synthesis of methymycin -  
 XX  
 XX Claim 19; Page 287-299; 438pp; English.  
 XX  
 CC The invention relates to an isolated and purified nucleic acid segment  
 CC comprising a desosamine biosynthetic gene cluster, a fragment or its  
 CC biologically active variant, where the nucleic acid sequence is not  
 CC derived from the eryC gene cluster of *Saccharopolyspora erythraea* or  
 CC streptomycetes antibioticus. The invention also relates to a macrolide  
 CC biosynthetic gene cluster, or fragments thereof. The macrolide  
 CC biosynthetic gene cluster encodes proteins which synthesise methymycin,  
 CC pikromycin, neomethymycin, narbomycin or a combination of these  
 CC compounds. Recombinant or augmented cells comprising the desosamine  
 CC and/or macrolide biosynthetic gene clusters are useful for the production  
 CC of biologically active macrolides. The macrolide biosynthetic proteins  
 CC are useful for synthesis of methymycin, pikromycin, neomethymycin and  
 CC narbomycin. The alternative termination of polyketide synthesis may be  
 CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)  
 CC monomers. The compounds produced by the recombinant host cells are useful  
 CC as biopolymers, e.g., in packaging or biomedical applications. To  
 CC engineer PHA monomer synthases or to prepare biologically active agents,  
 CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,  
 CC chronic obstructive pulmonary disease as well as other diseases involving  
 CC respiratory inflammation, cholesterol-lowering agents or macrolide-based  
 CC antibiotics which are active against a variety of organisms, e.g.,  
 CC bacteria, including multi-drug resistant pneumococci and other  
 CC respiratory pathogens, as well as viral parasitic pathogens, or as crop  
 CC protection agents (e.g., fungicides or insecticides) via expression of  
 CC polyketides in plants. The present sequence represents a protein  
 CC encoded by the desosamine biosynthetic gene cluster from *Streptomyces*  
 CC venezuelae ATCC 15439.  
 CC  
 XX  
 XX  
 SQ Sequence 3782 AA;  
 Query Match 21.7%; Score 951.5; DB 21; Length 3782;  
 Best Local Similarity 31.4%; Pred. No. 8.2e-75;  
 Matches 275; Conservative 123; Mismatches 340; Indels 137; Gaps 25;  
 QY 9 ILKRLTAEKVDLLAGIDFW-----HTRALPRHGVPSLFTDGPNGVGRKTFENG 58  
 DB 1121 LVAQMTELDEKISFY-----HMLADPDRONVGYLPGVPLRIGELPRARAGPNGIR-----LVG 1172  
 QY 59 VPACFPCTGSLGTFNQTLLEBAGKMGKEALAKSAHVILGPTINNOQSPILGREGFESI 118  
 DB 1173 QFATALPAPVALASTPFDTMADSGKYGVRGRLANDVILGPMNNIRVPHGRNVEYF 1232  
 QY 119 GEDEFLAGLAALIRIGTSGVQATIKHFLCDOEDRRMNOVSITYERALREIYALPQ 178  
 DB 1233 SEDPLVSSRTARVAQIKIGQAGLMTTAKHPRANNOENRNSVANNADDEQTLREIPEPAFE 1292  
 QY 179 IAVBDSQPGA---FMTAYNGINGVSCSENPBYDGLMRKEMGDLIWSDMYGYSTTEAV 236  
 DB 1293 AS---SKAGASAFSCAVANGNGKRSQCNDELNNVLTQWGFGQWVMSDLAT- PGDAI 1348

QY 237 VAGLDLE-----PGPP-REFGETLKFVNSNGK-PFIHVLDQRAREVLOQYKK 282  
 DB 1349 TKGLDQMGVELPGDVPKGPSPAKFEGEALKTAVLNGTVEAATVRSARIVGOMKEF 1408  
 QY 283 CAASGVTEGPEPTVNNPTETALALRKVNEGIVLLKNNVPLPS--KKKKTLLVGPNA 340  
 DB 1409 GLLATPAPRPE--RDKAGAOAVSRKVAENGAVLLRNNGQALPLAGDAGKSIAVIGPTA 1465  
 QY 341 KQATYHGGSAALRAYAVTPPDGLSKOLETPPSYTVGAVTVPPILGEOCLTPDGAPGM 400  
 DB 1466 VDPKVTGLSAHVPDPSAALPDLTIKAR-----AGACATVYETGETGTQIPAGN 1517  
 QY 401 RMRVFNPEPGTPNROHIDELFTKTDMLVDYHHPKADTWADMEGTYTADDCYTELG 460  
 DB 1518 LSPAFNG-----HOLE--PGKAGALY--DGLTTPADDEYRIA 1552  
 QY 461 LVVCGTAKAYVDDOLVVDNATKQVPGDAFPGSATRETRGRINLVKNTYKFKIEFGSAPT 520  
 DB 1553 VRATG--GYATVOL--GSHTEAGQYVGKVS--SPLKLTGK--THKL----- 1592  
 QY 521 YTLKGDITVPGHSLRVGGCKVIDDOAEIKESVALAKEHDOVITICAGLNADMETEGADRA 580  
 DB 1593 -TISGFAMSAITPLSELGWTTPAAADATIKAVESARKATAVVFA--YDGTGEGVDRP 1648  
 QY 581 SMKLPGLVDLIDVAAANPTVVVMTGTPEEMPMUDATPAVIAQAMYGNETGNSIADY 640  
 DB 1649 NLSTPGTODKLISAVADANPMTIIVLVNTGSSVLMPLSKTRAVLDMYPSQAGAEATPAL 1708  
 QY 641 VGDVNPBGKSLSPKRLQDNPAFLNRTFAG-----FTLGEDVYVGYRYEP 690  
 DB 1709 LKGDVNPBGKSLSPKRLQDNPAFLNRTFAG-----FTLGEDVYVGYRYEP 690  
 QY 691 ADKDVNPFPGHSLSTYTFEAFNSLVSH-KDKGLSVLSVKNVTSVPPAQVAAQLYKRLQA 749  
 DB 1762 ENVKPLFEPGHSLSTYTFEAFNSLVSH-KDKGLSVLSVKNVTSVPPAQVAAQLYKRLQA 749  
 QY 750 AKINRPFVKELKFAVELQPGETKAVTIEDEK-----YVAAAFDE 790  
 DB 1822 VTAPQAKKRLVGYTVSLAAGAKTVYVNDRLQLOTGSSSADLRGSAITVVMSSRAET 1881  
 QY 791 ERDQMCVERGDEVIVSDSSAAKDVALLRGFTVG 825  
 DB 1882 PRVPLDILKAAVELRAETDAIARVLDSGRYLLG 1916  
 RESULT 5  
 ID AAE24228 standard; Protein; 3782 AA.  
 AC AAE24228;  
 DT 23-SEP-2002 (first entry)  
 DE Streptomyces venezuelae desosamine gene cluster encoded protein #1.  
 KW glycosylated polyketide; modified recombinant bacterial host cell;  
 KW mBHC; macrolide; anthracycline; angucycline; avermectin; milbemycin;  
 KW tetracycline; polyene; polyether; ansamycin; isochromanone; sugar;  
 KW desosamine; des.  
 OS Streptomyces venezuelae.  
 PN WO200229035-A2.  
 PD 11-APR-2002.  
 PE 05-OCT-2001; 2001MO-US31255.  
 PR 05-OCT-2000; 2000US-238185P.  
 PA (MIND) UNIV MINNESOTA.  
 PA (LITUH) LITU H.  
 PA (SHER) SHERMAN D H.

PA (ZHAO/) ZHAO L.  
 XX Liu H, Sherman DH, Zhao L;  
 PT WPI; 2002-405171/43.  
 DR  
 XX  
 PT Modified recombinant bacterial host cells in which the expression and  
 PT activity of nucleic acids encoding sugar biosynthetic enzymes has been  
 PT altered, useful for producing metabolites with altered sugar structures  
 PS  
 XX  
 PS Disclosure; Fig 8; 1/4pp; English.  
 CC  
 CC The invention provides a method to alter the sugar structure diversity  
 CC for a particular metabolite via the recruitment and collaborative action  
 CC of sugar genes from a variety of sugar biosynthetic pathways to yield a  
 CC metabolite comprising a non-natural sugar, e.g., a novel glycosylated  
 CC polyketide. The invention also relates to a modified recombinant  
 CC bacterial host cell (mBHC) in which the expression and activity of  
 CC nucleic acids encoding sugar biosynthetic enzymes has been altered.  
 CC The mBHCs may be cultured to produce the modified sugar products,  
 CC e.g., a macrolide, anthracycline, angucycline, avermectin, milbemycin,  
 CC tetracycline, polyene, polyether, ansamycin or isochromanone.  
 CC The present sequence is Streptomyces venezuelae sugar (desosamine)  
 CC biosynthetic gene cluster encoded protein.  
 CC  
 SO Sequence 3782 AA;  
 Query Match 21.7%; Score 951.5; DB 23; Length 3782;  
 Best Local Similarity 31.44; Pred. No. 8.2e-75;  
 Matches 275; Conservative 123; Mismatches 340; Indels 137; Gaps 25;  
 QY 9 ILKLTLEAKYDLAGIDFW-----HTKALPKHVPSSLFTGPGVCGTKFENG 58  
 DB 1121 LVAQMTLDEKISFY---HMLADPDQNVGILPGVPRGRIGTELAAGSPKIR---LVG 1172  
 QY 59 VRAACEPGCTSGSPFNQTLLEEGKMKMGKFAIKSAHVIILPTINMQRSLGREGESI 118  
 DB 1173 QYATAPAPVALASTFDDTMADSYKVGKRGKRALNDQMTVGLPMNNINRVPHGRNTEF 1232  
 QY 119 GEDPFLAGLGAALIRGIQSTGVQATIKHFLCNDQEDRRMVOISVTERALREYALPFO 178  
 DB 1233 SEDPLVSSRTVAOAKGIGAGIMTAKHFAANNENRFSVNNVDEQTLRETFEAPFE 1292  
 QY 179 IAVRDSQPCA--PMTAINGINGVSCSENPKYLDGKLREKMGDGLINSQWYGYSTTEAV 236  
 DB 1293 AS---SKAGAASPMCAVYNGLNKSPSCGNDELNNVLTQMGFGQVWMDLALAT-PGIDAI 1348  
 QY 237 VAGLDLE-----PGPP-REFGETLKFVNSNGK-PFIHVLDQRAREVLOQYKK 282  
 DB 1349 TKGLDQMGVELPGDVPKGPSPAKFEGEALKTAVLNGTVEAATVRSARIVGOMKEF 1408  
 QY 283 CAASGVTEGPEPTVNNPTETALALRKVNEGIVLLKNNVPLPS--KKKKTLLVGPNA 340  
 DB 1409 GLLATPAPRPE--RDKAGAOAVSRKVAENGAVLLRNNGQALPLAGDAGKSIAVIGPTA 1465  
 QY 341 KQATYHGGSAALRAYAVTPPDGLSKOLETPPSYTVGAVTVPPILGEOCLTPDGAPGM 400  
 DB 1466 VDPKVTGLSAHVPDPSAALPDLTIKAR-----AGACATVYETGETGTQIPAGN 1517  
 QY 401 RMRVFNPEPGTPNROHIDELFTKTDMLVDYHHPKADTWADMEGTYTADDCYTELG 460  
 DB 1518 LSPAFNG-----HOLE--PGKAGALY--DGLTTPADDEYRIA 1552  
 QY 461 LVVCGTAKAYVDDOLVVDNATKQVPGDAFPGSATRETRGRINLVKNTYKFKIEFGSAPT 520  
 DB 1553 VRATG--GYATVOL--GSHTEAGQYVGKVS--SPLKLTGK--THKL----- 1592  
 QY 521 YTLKGDITVPGHSLRVGGCKVIDDOAEIKESVALAKEHDOVITICAGLNADMETEGADRA 580  
 DB 1593 -TISGFAMSAITPLSELGWTTPAAADATIKAVESARKATAVVFA--YDGTGEGVDRP 1648  
 QY 581 SMKLPGLVDLIDVAAANPTVVVMTGTPEEMPMUDATPAVIAQAMYGNETGNSIADY 640

```

Db 1649 NLSLPGDOKLISAVADAMPNTIVLNTGSSVLMPLMSTRVLDWMFGAGAEATRAL 1708
QY 641 VFGDVPDSKLSLSPFKLQDNPAPLNFRTENG-----RTLGEDEVYGYRYEYF 690
Db 1709 LYDVAVPSGLTQSF-----PAAENOHAAVAGDPTSPGVDMQOTYREGJHVGCRWFDK 1761
QY 691 ADKDVNPFPGHGLSTYTFEAFNLSVSH-KDGKLSVLSVKNKGSVPQAVOLVVKPLQA 749
Db 1762 ENVKPLPEFGHGLSTSTFQSNAPTVVRSTGGLKATYTVNRKSGRAGOEVOATLGASPN 1821
QY 750 AKINRVRKELGFAKVELQPGCTKAVTIEBOK-----YVAAYFDE 790
Db 1822 VYAPQAKKLVGTYTVSLAAGAKTVTVNDRQLQTGSSADLRGSAATVVMWSRAET 1881
QY 791 ERDQMCVEKGEYIVSDSAKGDVALRGKFTVG 825
Db 1882 PRVPFLDLKAAVEELRAETDAIARVLDGSKRLG 1916

```

## RESULT 6

AAV77189

ID AAV77189 standard; Protein; 808 AA.

AAV77189; 05-JUN-2000 (first entry)

S. venezuelae macrolide beta-glycosidase Desr, SEQ ID NO:24.

Desosamine biosynthesis: macrolide; polyketide; methymycin; pikromycin;  
 neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;  
 biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;  
 chronic obstructive pulmonary disease; respiratory inflammation;  
 hypercholesterolemia; crop protection agent.

Streptomyces venezuelae ATCC15439.

MO200000620-A2.

06-JAN-2000.

25-JUN-1999; 99MO-US14398.

26-JUN-1998; 98US-0105537.

(MINU) UNIV MINNESOTA.

PI Sherman DH, Liu H, Xue Y, Zhao L;

DR WPI: 2000-160679/14.

N-PSDB: AAZ87294.

Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.

synthesis of methymycin and pikromycin.

Claim 19; Page 371-373; 438pp; English.

The invention relates to an isolated and purified nucleic acid segment  
 comprising a desosamine biosynthetic gene cluster, a fragment or its  
 biologically active variant, where the nucleic acid sequence is not  
 derived from the eryc gene cluster of *Saccharopolyspora erythraea* or  
 streptomycetes antibiotics. The invention also relates to a macrolide  
 biosynthetic gene cluster, or fragments thereof. The macrolide  
 biosynthetic gene cluster encodes proteins which synthesise methymycin,  
 pikromycin, neomethymycin, narbomycin or a combination of these  
 compounds. Recombinant or augmented cells comprising the desosamine  
 and/or macrolide biosynthetic gene clusters are useful for the  
 production of biologically active macrolides. The macrolide biosynthetic  
 proteins are useful for synthesis of methymycin, pikromycin,  
 neomethymycin and narbomycin. The alternative termination of polyketide  
 synthesis may be useful to prepare novel antibiotics and  
 polyhydroxyalkanoate (PHA) monomers. The compounds produced by the

recombinant host cells are useful as biopolymers, e.g., in packaging or  
 biomedical applications, to engineer PHA monomer synthases, or to prepare  
 biologically active agents, such as chemotherapeutics,  
 immunosuppressants, agents to treat asthma, chronic obstructive pulmonary  
 disease as well as other diseases involving respiratory inflammation,  
 cholesteral-lowering agents or macrolide-based antibiotics which are  
 active against a variety of organisms, e.g., bacteria, including  
 multi-drug resistant pneumococci and other respiratory pathogens, as well  
 as viral parasitic pathogens, or as crop protection agents (e.g.,  
 fungicides or insecticides) via expression of polyketides in plants.  
 Sequences AAV77181-777189 represent desosamine biosynthetic enzymes from  
 Streptomyces venezuelae ATCC 15439, which are encoded by sequences  
 AAZ87286-287294.

SQ Sequence 808 AA;

Query Match 21.5%; Score 944.5; DB 21; Length 808;

Best Local Similarity 32.2%; Pred. NO. 2.6e-75;

Matches 275; Conservative 112; Mismatches 325; Indels 143; Gaps 26;

```

QY 9 ILKLTTLAEKVDLAGIDFW-----HTKALPKHGVPSLFTDGPNGVGRFENG 58
Db 55 LVAQMTLDEKISFY-----HMLDPRONVGYLPVPRGLGPELRADGPGIR----LVG 106
QY 59 VPAACFGCTSLGSPNPTLLEAGKMKKEALAKSAHVLTGPTIMKNSPLGREGFESI 118
Db 107 QTATLAPAPVALASTFDOTMADSYGKMGKRDGLALNDMDVGLPMNNIRVPHGSRNRYEYF 166
QY 119 GEDPFLAGLGAALIRIGISTGVQATIKHFLKNDQEDRRMVOISYTERALREYALPRQ 178
Db 167 SEDPLVSRFAVQIKIGQAGLMTTAKHFAANNQNNNSVAVNDEOTLRIEPAFE 226
QY 179 IAVVDSQPGA--FMTATANGINGVSCSNPKYLDGMLRKEMGDLTMSDMYSTTEAV 236
Db 227 AS---SKAGASFCAYNGALNGKRPSCGNDELNNVLRTOGMGCGWMSDMLAT--PGTDAI 282
QY 237 VAGIDLEM-----PGPP--RFRGETLKFNYSNGK--PEIHVDQRAEYVQPKK 282
Db 283 TKGLDQEMGEVELGDVPEKGPSPKAFEGEALKTAVLNGTVEPAVTSAEIVQOMERF 342
QY 283 CAASGVTEGPEPTVNTPTAALLRKVGNEGIVLLKNNVPLPS--KKKTLIVGPA 340
Db 343 GLLATPAPRPE---RDKAGAQVSRKVAENGAVILLRNQGLPLAGDAGKSIAVIGPTA 399
QY 341 KQATYHGGSGAALRAYVATPPEDGLSKOLETPPSTVGAITYVPTIIBOCITPGAGGM 400
Db 400 VDKVYIGLSAHVPPSAAPLDTIKAR-----AGCATVYETEGETFGTOIPAGN 451
QY 401 RMRVNEPPEPTPNRQIHIDELFTKTDMLVDYHPRKADVTWADMEGTVTADDECTYELG 460
Db 452 LSPAFMNG-----HGLE---PGKKGALY---DGLTVPAPAGEYRIA 486
QY 461 LVVCGTAAKAYVDQLVVDNATKQVGDAPFGSATREERGNLNVKNTYKRIEFGSAPT 520
Db 487 VRAATG--GTAIVQL-----GSHFTIAGOVYKVS---SPLKLTGK--THKL----- 526
QY 521 YTLKGDPIVPGHGLSVGCKVVIDQAEITEKSVALAKHEDQVITICAGLNAMMETEGADRA 580
Db 527 -TISGFAMSAATPLSLLEGVWTPAADAATIAAVESAKARAPVFA---YDGESEGDV 562
QY 581 SKKLEPVLDQLIADYAAANPNTVVVMTGTPEEMPMILDATPAVIOAYGNGENSTADV 640
Db 583 NLSLPGTOKLISAVADAMPNTIVLNTGSSVLMPLMSTRVLDWMYPOAGAEATRAL 642
QY 641 VFGDVPDSKLSLSPFKLQDNPAPLNFRTAG-----RTLGEDEVYGYRYEYF 690
Db 643 LVGDVNPSEKLTQSF-----PAAENOHAAVAGDPTSPGVDMQOTYREGJHVGCRWFDK 695
QY 691 ADKDVNPFPGHGLSTYTFEAFNLSVSH-KDGKLSVLSVKNKGSVPQAVOLVVKPLQA 749
Db 696 ENVKPLPEFGHGLSTSTFQSNAPTVVRSTGGLKATYTVNRKSGRAGOEVOATLGASPN 755
QY 750 AKINRVRKELGFAKVELQPGCTKAVTIEBOKYVAAYFDEERDQMCVEKGEYIVSDS 809

```

Db	756	VTAQAQAKRLVGTYSIAGAEKVTYTVN-----DRQLDTGS	794
Oy	810	SAAMDVALRGKFTV	824
		:	
Db	795	SSAD---LRGSAIV	805
RESULT 7			
ID	AAE24237	standard; Protein; 809 AA.	
XX	AC	AAE24237;	
XX	DE	23-SEP-2002 (first entry)	
XX	XX	Streptomyces venezuelae DesR gene encoded protein.	
KW	Glycosylated polyketide; modified recombinant bacterial host cell;		
KM	mRBHC; macrolide; anthracycline; angucycline; avermectin; milbemycin;		
KW	tetracycline; polyene; polyether; ansamycin; isochromanonequinone; sugar;		
KM	desosamine; DesR; glucosidase; enzyme.		
XX	OS	Streptomyces venezuelae.	
FH	Key	Location/Qualifiers	
FT	Misc-difference 1	/note= "Encoded by GTG"	
PN	MO200229035-A2.		
PD	11-APR-2002.		
XX	05-OCT-2001; 2001WO-US31255.		
Pf	05-OCT-2000; 2000US-238185P.		
XX	(MINU ) UNIV MINNESOTA.		
PA	(LIUH/) LIU H.		
PA	(SHER/) SHERMAN D H.		
PA	(ZHAO/) ZHAO L.		
PI	Liu H, Sherman DH, Zhao L;		
XX	WPI; 2002-405171/43.		
DR	N-PSDB; AAD39043, AAD39052.		
PT	Modified recombinant bacterial host cells in which the expression and		
PT	activity of nucleic acids encoding sugar biosynthetic enzymes has been		
PT	altered, useful for producing metabolites with altered sugar structures		
Disclosure; Page 171-173; 174pp; English.			
The invention provides a method to alter the sugar structure diversity for a particular metabolite via the recruitment and collaborative action of sugar genes from a variety of sugar biosynthetic pathways to yield a metabolite comprising a non-natural sugar, e.g., a novel glycosylated polyketide. The invention also relates to a modified recombinant bacterial host cell (mRBHC) in which the expression and activity of nucleic acids encoding sugar biosynthetic enzymes has been altered. The mRBHCs may be cultured to produce the modified sugar products, e.g. a macroide, anthracycline, angucycline, avermectin, milbemycin, tetracycline, polyene, polyether, ansamycin or isochromanone. The present sequence is Streptomyces venezuelae sugar (desosamine) biosynthetic gene cluster DesR (glucosidase) gene encoded protein.			
SQ	Sequence	809 AA:	
Query Match	21.5%;	Score 944.5;	DB 23; Length 809;
Best Local Similarity	32.2%;	Pred. No. 2.6e-75;	
Matches 275; Conservative	112;	Indels 143;	Gaps 26;

OY		9	I LKLTLLAEKVVDLLAGIDFW-----HTKALPKRGVSLFETPGPNCNGRTEFFNS	58
Db		56	LVAOMTDEKISFEV---HMAIDPDROUNGYLPYGPBRIGTEBELRAADOPNCR---	107
OY		59	VPAACFCGCGSISTSTENOITLLEAGCKMKGEAIAVASHVILGPTIINMORSPLGREGEST	118
Db		108	QTATALLPPVALASTEDPDITADSYGVNMORDORALNDQWVGPMNNIRHPGRGRNYETE	167
OY		119	GEDPFGLAGLGAALINGIOSTGYOATIKFHFLCNDQEDRRMWOSIVTERALRETYALPFQ	178
Db		168	SEDDPVSSRTVAQIKGIQOAGMLTTAKHEFAANNQDNENNFESVANAVDQTLRELIEPAFE	227
OY		179	IAYRDSOGCA--EMFTAVNGINGVSCSNPEYLDGMRLKEEMGDGLMSDWYGTYSTBAV	236
Db		228	AS---SKGAAEFMCANGLNGKPRSCGNDBDLLNNVLRTQMFGQWMVSDMLAT- PGTDAI	283
OY		237	VAGLDLEM-----PGDP-RFRGETLKFNYSNK - PEIHVIQBARAREYLQFYKK	282
Db		284	TGDIQOENGVELPGDVPKGEPSPAPKFFGALKTAALNTGVBDAVTSKAERIVQOMKF	343
OY		283	CASGVTVENGPEPTYNNNPETAALLRYVNEGIVLTKKENNVLPJS - KKKKTLLVFNPA	340
Db		344	GILLATPADRP---RDKAQAQVSRKVAENMGALILRNCOGALPLAGDGKSAIVIGPTA	400
OY		341	KQATYHGGSAAIRAYAVTPFDGLSKOLETPPSYTGATTYPPIILSGOCCLPPGAGCM	400
Db		401	VDPKVTGIGLSAHVVDSSAAPLDTIKAR-----AGAGATVYTGETGETGTOTIPAGN	452
OY		401	RMEVNEPERPCTNRNHIDEFLTUKDMLVYHYPRKAADPWADVAMEGYTATBEDCTYELG	460
Db		453	LSBAPFOG-----HQLE--PGKAGALY--DOTLVPADGETRPA	487
OY		461	LYVCGTAKAYVDDOLVYNATKOVPDGFPGSATPRETGRI NLYKGWTYFKILEFSAPT	520
Db		488	VRATG---GYATVOL---GSHTEIAGOVYKVS-- SPLTKILKG-THLK-----	527
OY		521	YTLKGTIYPBGHSILAVGCCVKYIDDOAFIEKSVALAKEHDVILICAGINDMETEGADRA	580
Db		528	-TTISGFMSATPLSELGWVTPAADATITAKAVEABRKARFAYVFA-- YDDGTGVDPR	583
OY		581	SMRLPGVILDLIDVAANPNNTVVVMQGTPEEMFIDLATPAVLOANYGNGEFGNSIADV	640
Db		584	NLSLRPGODKLISAVDAMPNTIVLVNTSSVLMPLSKTRAVLDMMTPGOGAGEATIAL	643
OY		641	VFGDYNPFGKLSISFPKRLODNPAPLFNFTENG-----PTLYGEDVYVSYRYEEF	690
Db		644	LYGVNPNPFGKLTQSF-----PAENQNAVAGDPTSPGVNDNQOTYREGIHVEYRMFDK	696
OY		691	ADKDVPNFPFGHLSYTFPAFNLVSZH-KDGKLSVLSKYNTGSTPGAQVAMQLVYKRLQA	749
Db		697	ENVKPLDFPFGHLSYTSFTQSAPTVRTSTGGILKIVTVYRNGKRAQGOEVQATLGASP	756
OY		750	AKINPVEKLKGFAKVELQGETKAVTIEQEKVYAAYFERDERDQCWEKGDYEYIVSDS	809
Db		757	VTAPPAKKRLLGVYKRVSLAAGEARTVNV-----DRROLQGCS	795
OY		810	SAAKDGVALRGKETVY	824
Db		796	SSAD----LRGSATV	806
RESULT		8		
AAM34558	ID	AAM34558	standard; Protein; 721 AA.	
XX		AAM34558;		
XX		12-MAR-1998	(first entry)	
XX			Thermotoga maritima MS86 glycosidase.	
XX			Glycosidase; thermostable; textile; food processing; pharmaceutical;	
KW			detergent; baking; industry; Thermococcus; Staphylothermus;	

KW Pyrococcus; glucose; soluble oligosaccharide.  
 OS Thermotoga maritima.  
 XX WO925417-A1.  
 XX  
 XX 17-JUL-1997.  
 PD  
 XX 10-JAN-1997; 97WO-US00092.  
 PF  
 XX 13-SEP-1996; 96US-0712612.  
 PR 11-JAN-1996; 96US-0583787.  
 XX  
 XX (RECO-) RECOMBINANT BIOCATALYSIS INC.  
 PA  
 XX Bylina EJ, Lam DE, Mathur EJ, Swanson RV;  
 PI WPI: 1997-372858/34.  
 DR N-PSDB; AAT93682.  
 XX  
 PT New thermostable glycosidase(s) - from Thermococcus, Staphylothermus  
 PT and Pyrococcus, used in the textile, food processing,  
 PT pharmaceutical, detergent and baking industries  
 PT  
 XX Claim 4; Fig 5; 82pp; English.  
 XX  
 CC The present sequence represents glycosidase isolated from Thermotoga  
 CC maritima. The enzyme or its encoding nucleic acid sequence is used for  
 CC generating glucose from soluble oligosaccharides. The enzyme can be  
 CC used in the food processing, pharmaceutical, textile, detergent and  
 CC baking industries. The enzyme is also used to treat lactose intolerance,  
 CC as a diagnostic reporter molecule, in corn wet milling or in the fruit  
 CC juice industry. The enzymes can be used to hydrolyse guar gum to remove  
 CC non-reducing terminal mannose residues. The nucleic acids encoding the  
 CC enzyme may be used to generate probes to identify similar sequences.  
 CC  
 SO Sequence 721 AA:  
 Query Match 20.7%; Score 907; DB 18; Length 721;  
 Best Local Similarity 30.4%; Pred. No. 5, 1e-72;  
 Matches 263; Conservative 121; Mismatches 286; Indels 194; Gaps 29;  
 QY 6 VEAILKTLTAKVLLAGIDF-----WHTKALPKHGVPSLRTDGPNGV 50  
 DB 4 IDEILSQTTEKVKLVVGVGLPGLEGNPHSRVAGAAGETHPVRGLGPAFVLADGAGL 63  
 QY 51 R-----GTRFENGVPACPCGTSIGSTFNCOTLEAGKMGKEAIASAVIIGPTINMO 106  
 DB 64 RINPTRENDENTYTTTAPVEIMLASTWNRDLLEVGKANGEEVREYGVVLLAPAMNH 123  
 QY 107 RSLGGRGEESIGDEPPLAGIAALRGIOSTGVQATIKHFLCNDQEDRRMAYOSIVTE 166  
 DB 124 RNPICGRFEYYSDEPVALSGEMASAFVKGVOGACIHFVANNQETNRVVDIIVSE 183  
 QY 167 RALREIYALPPQIAVRSOPGFMTAVNGINGVSCSENPXYLLDGMARKENGMDGLMSDW 226  
 DB 184 RALREIYALGPFEIYAKARPMYMSAKNKGATCSQENELKKVLEEMGRGGEVMSDW 243  
 QY 227 YGTSITEAVVAGLDLMPGP-----RFEKTELFVNSGKPFPIHYIDORAREVLOF 279  
 DB 244 YAGNPEYQILKAGMDIMPGRAYQVNTERRDEIEETMEALKEGKISEVLDLDCVNIILK- 302  
 QY 280 VKKCAASGVTENGPE---TVNNTPEP-----AALLKRVNGEGIVLLKENNNVPLSKKKKT 333  
 DB 303 -----VLVNAPEFGYRYSKKPLESHAEVAYEAGAEVLLLEN-NGVLPDENHIV 353  
 QY 334 LIYGPNAKQATYHGGSAALRAYAVTPDGLSKOLETPSYTGAVYTVPIIGEOGLT 393  
 DB 354 AVFGTQIEIKINGTSG-----DHPRTTISIEFI----- 384  
 QY 394 PDGAPGRMRVFNPEPGTNRQHIDELFTTKTDMHLDVYHFKAADTYVADMEGTYADE 453  
 DB 385 -KGTG-----SGDTHPRYTISEEYIKK--MKRETEYKPR-TDSW----- 420

QY 454 DCTVELGLVCGTAKAYVDQLVYDNATKQVPGSAFEGSARTRESTGINIVKNTYKFI 513  
 DB 421 -----GTVI-----KPKLPENFISEKETKKPKKNDAVAVV-----ISRISEGYDRK- 463  
 QY 514 EFGSAPTYTLKGDITVDPHGSLRVGGCKVIDDQAEIKSVYALAKE-HDQVYICASLMDW 572  
 DB 464 -----PVKGFDEYLS-----DDELELIKTV--SKEPHDO----- 489  
 QY 573 ETEGADRASKMLPGVLDQLIADVAANPNVVMQGTPEEM-PWLDATPAYIQAYGCGN 631  
 DB 490 -----GKVVVLLNMGSPLEVASMRDLVDGLILVWQAOQ 523  
 QY 632 ETGSIADVVFEGDYNPSGKLSLSEPKRLQDPNPAFLNFTPEAGR--TLYGEDV-----YVG 684  
 DB 524 EMGRIVADVGVKINPSGKLPTEFPKDYSDVPSW-----TFPKKLPTEFPKDYSDVPSWTF 579  
 QY 685 YRYEFADKDVNPEPGHGSYTFEAFSNLSVSHKDKLSVLSVKNMGSVPGAQVADLVV 744  
 DB 580 PRYDTGTGVPAYEFGISLYTEFEYKDIKIDGETTLRVSYTINTGDFRAGKEVSYI 639  
 QY 745 KPLQAKINRPVKELGFAKVE-LQPGETKAVTIEQEKYVAAYDEDERDQVCZKGYE 803  
 DB 640 K-APKGKIDKPFQELKAFKTKLNGESELSTELPLDLASFQDK- --WVESGEYE 695  
 QY 804 VIYSDSAAKDGVALGKFTV-GE 826  
 DB 696 VRVGASSR--DIRLNDIFLVEGE 716  
 DE  
 XX 18-OCT-1996 (first entry)  
 XX  
 DE Chinaeric thermostable beta-glucosidase.  
 XX  
 KW Chimaera; beta-glucosidase; Cellvibrio gilvus; Agrobacterium tumefaciens;  
 KW homology; decomposed cellulose; glucose; cellobiose; cellobiose;  
 KW cellulotriase; cellobiose.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Region /note- "amino acids derived from Cellvibrio gilvus  
 FT 693..752  
 FT /note- "amino acids derived from Agrobacterium  
 FT tumefaciens beta-glucosidase residues 759-818"  
 JP08131168-A.  
 XX  
 PD 28-MAY-1996.  
 PF  
 XX 09-NOV-1994; 94JP-0299049.  
 PR  
 XX 09-NOV-1994; 94JP-0299049.  
 (NORQ) NORINSUISANSO SHOKUHIN SOGO.  
 WPI: 1996-311636/32.  
 DR N-PSDB; AAT32999.  
 PT  
 PT Thermostable chimaeric beta-glucosidase - useful for prodn. of  
 PT cellobiose from decomposed cellulose  
 XX  
 PS Claim 2; Page 5-8; 9pp; Japanese.  
 CC This is the amino acid sequence of a novel chimaeric beta-glucosidase



CC enzyme composed of amino acids 1-692 of *Cellulibrio gillvus*  
 CC beta-glucosidase and amino acids 759-818 from the *Agrobacterium*  
 CC *tumefaciens* beta-glucosidase, replacing amino acids 693-752 of the  
 CC *Cellulibrio* sequence. The chimera was constructed by isolating the  
 CC *C.gillvus* beta-glucosidase gene and comparing its sequence with the  
 CC *ruminococcus albus*, *butyrivibrio fibrisomens*, *agrobacterium tumefaciens*,  
 CC *corresp.* genes from several other species, e.g. *Hansenula anomala*,  
 CC etc. The homologous sequences from the other species could then be used  
 CC to generate chimeric glucosidase genes. The novel chimeric protein  
 CC acts on decomposed cellulose to generate glucose and cellobiose, and has  
 CC a reduced ability to decompose cellobiose as compared to *cellulotriose*,  
 CC *cellobiose* or *cellopentase* as substrate. The novel enzyme has an  
 CC optimum pH of 6.0 and temp of 40-45 deg. C, is stable in pH 4-9 and at  
 CC 35 deg. C.

XX Sequence 752 AA;

Query Match 19.8%; Score 869; DB 17; Length 752;  
 Best Local Similarity 29.3%; Pred. No. 1.4e-68;  
 Matches 250; Conservative 107; Mismatches 279; Indels 220; Gaps 22;

OY 2 ADIVVEALIKKTLAEKVDLAGI--DFW---HTKALPKH-----GVPSLRFT 44  
 DB 51 ADQARARLAVQAMTQGEKLRWVFYGFHDFGSKKHDPALQSGAYIPGPRRLGALPAPFT 110  
 OY 45 DGPNGVGTGTFNGVPAACPCGTSLSSTFNQTLLEAGKMGKALAKSAHVILGPTIN 104  
 DB 111 DAGOGVASQSGANREKALPESGLSTASTWDPKVAAYAGMISEKARASGFENWMLAGVNA 170  
 OY 105 MORSPLGGRGESIGEDPFLAGLGAALIRGISTGVQATIKHFLINDQEDRRMAYOSIV 164  
 DB 171 LQREPRNGRNFYAGEPDLPLAGTIGALIGVESNRILSTLKHVILNDQETGNEIDARI 230  
 OY 165 TERALREIYALPFOIYAVDSQAFMTAYNGINGVSCSENPKEIYDGLKREKMDGLINS 224  
 DB 231 DKAALRMSDLLAMELALBQSDAGSMAYNRLNGPYCEHFWLLSEFLKDWGFRGVMS 290  
 OY 225 DWYGTYSSTTEAVVAGLD---LEMPGPRRGFTLKFNVSNGKPFIVHIDQARE--VLD 278  
 DB 291 DWGATHSTVAANSGLDQSGOEDPKSYF--GGALIEAVKVG-----AVQKRLDDWVTR 344  
 OY 279 FVKKCAASGVTEGNETTV-----NNTPEALLRKVGNEGIVLLKMDNNVPLSK 330  
 DB 345 IYRTMFGKGVVNDPLKPGALIDFAANGQPPD-----GEEGMVLLKMGGRLLPLAKT 396  
 OY 331 KKTLL-IVGNKQATYHGGGSAALRAYAVTPFDLSKQLETTPSYVAVTTPILGE 389  
 DB 397 VRTIANVIGHADAGVLSGGSSQV-----YFVG-----IAV 428  
 OY 390 OCCTPGAGRMRYFNEPPTPNRQIHDELFTKTDMLVDYHYHKAADTWADMEGTY 449  
 DB 429 KGLLPRTWVG-----PYYVYPSPLRAIOAOPNA----- 458  
 OY 450 TADECTYELGLVCGTAKAYVDOLVVDNATKQVPGAFSGATRETGINLVKNTY 509  
 DB 459 ----- 458  
 OY 510 KEKIEGSAPTYTLKGDITVPGHSLRVGCKVIDDOAEIEKSVALAKEMDOVILCAGLN 569  
 DB 459 --KV-----VFDDGRDPRARAANAAYAGAVVALVEAN-- 486  
 OY 570 ADMETEGADRAASKLPGVLDOLIAVAAANPNTVVVMOTGPPEEMPIIDATPAYIQANYG 629  
 DB 487 -QWIGEMNADQTLALPDGQELITTSVAGANGRTVVVLQGGFVTPMPLARPAVLEAWP 545  
 OY 630 GNETGSIADVYVGDYNPSSKLSLSPKRLQD-----NPALFNRTAEAGRTLYG 678  
 DB 546 GTSQGEAIANVILGAVNPSGHLPATFPQSEQLPPKKLDGDKNPE-LQFAND-----YH 599  
 OY 679 EDVYVGRYRYEADKDVNPFPGHGLSYTTFPAFNSLSHSDKLSVSLVKNTGVPQAQ 738  
 DB 600 EGAAYVGIKWDLKGHRPLPFPGHGLSYTTFPAFNSLSGOLKDGRLHVRKVTNTGVNAGKD 659

OY 739 VAQLYVKPIQAAKINRPVKELGFAKVELQPGETAKVITEQEKVA----ATPEERDQ 794  
 DB 660 VPQVYAAPM-STFWMAP-KRLAMSKVALLPGETRGATGTAVLKTRADLAPFVDEGR 717  
 OY 795 WCVEKGDYEIVYSDSS 810  
 DB 718 FRADAGKYEILVANA 733

# RESULT 10

AAW69761  
 ID AAW69761 standard; Protein; 735 AA.

AAW69761;

DT 23-NOV-1998 (first entry)

DE Acetobacter xylinum beta-glucosidase.

XX Acetobacter xylinum; sucrofermentans; cellulose synthesis complex;  
 KW bcsA; bcsB; bcsC; bcsD; CMCase; beta-glucosidase; enzyme; cellulose;  
 KW microorganism.

XX Acetobacter xylinum.

OS W09839455-A1.

XX PD 11-SEP-1998.

PF 09-OCT-1997; 97NO-JP03633.

XX 04-MAR-1997; 97JP-0063927.

XX (BIO-P) BIO-POLYMER RES CO LTD.

PA Hayashi T, Tahara N, Tonouchi N, Tsuchida T, Yano H;

PI Yoshinaga F;

DR WPI; 1998-495854/42.

XX N-PSDB; AAV52831.

PT Gene encoding Acetobacter xylinum cellulose synthetase complex -  
 PT containing a group of genes including those for conventional and  
 PT novel "b-glucosidases

PS Claim 8; Pages 36-37; 50pp; Japanese.

XX This represents the amino acid sequence of a Acetobacter xylinum

CC subspecies sucrofermentans beta-glucosidase. The invention provides  
 CC a gene encoding a Acetobacter xylinum subspecies sucrofermentans derived  
 CC cellulose synthesis complex-produced protein. The gene sequence  
 CC represents bcsA, bcsB, bcsC or bcsD, CMCase and a beta-glucosidase  
 CC encoding gene. The novel gene and the enzyme participate in the synthesis  
 CC of cellulose by microorganisms. Cells transformed with the genes may be  
 CC used in the production of cellulose.

XX Sequence 735 AA;

Query Match 19.6%; Score 860; DB 19; Length 735;  
 Best Local Similarity 28.3%; Pred. No. 8.8e-68;  
 Matches 241; Conservative 125; Mismatches 277; Indels 210; Gaps 22;

OY 2 ADIVVEALIKKTLAEKVDLAGID-----FWHTKALPKHGVPSLRFTD 45  
 DB 37 ADARARQVLAASKLEKMSLTSYVDOGGFNGSVAPPGGSAAYLRAPQSGSLPQLQISD 96  
 OY 46 GPNVGRTGTF--NGVPAACPCGTSLSSTFNQTLLEAGKMGKALAKSAHVILGPTI 103  
 DB 97 AGCGVRNPAHRRNG--EAVSLPSGOSTASTWMDMARQAGVAMIGRAMQSGNILLGGA 155  
 OY 104 MORSPLGGRGESIEDPFLAGLGAALIRGISTGVQATIKHFLCNDQEDRRMAYOSI 163  
 DB 156 DLTRDPRGRNFEYAGBDPLQTRMVGSTTAGVQSOCHVISTIKHAMNDLETSRMTMSAD 215



Db 508 -----YPPPEVCLVETKLTWARGCTDKRLSTENDMNSTAYVNNARACPTITVYTHSG 558

QY 610 TPPEMFLD--ATPAVIOAWYGNENGTNSIADVFGDYNPSGLSLSPFKRLD--NPAEL 666

Db 559 GINTMRADNANTALIAHYPQENGNSIMDLIEDVAPSGRLPTIKLATDYCFPV 618

QY 667 NFRTEA----GRTLYGEDVYVGRYERADKDVNFPFGHGLSYTF 708

Db 619 NITNEADDPYVMOADFEGLIDIRHFDARNITPLEYEFGLSYTFEIEGVANLVAKSA 678

QY 709 ---AF--SNLSVSHKDK-----LSVLSVKNKTSVPGAOVQOLYVK--PLQAKINRPV 756

Db 679 KLSAFPASSTDISHPGPDIMEEVSYTAANKTGSVSGVQVQLISLPADGIPENSFM 738

QY 757 KELGFAKVELQPGETKAVTIEDEKVAAYFDEERDQWCEKDYEVIVSDSS 810

Db 739 QVLRGFKVDLPQGSQSVFESIMRDL--SFMTTADQMEIPNGQLEFRVGFSS 791

RESULT 12

AAW35004

AAW35004 standard; Protein; 778 AA.

AC AAW35004;

DT 21-MAY-1998 (first entry)

DE Thermotoga maritima endoglucanase.

KW Endoglucanase; cellulase; carboxymethylcellulose; cellulose;

KM biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;

KW thermostable enzyme; thermophilic; glycosidase.

OS Thermotoga maritima strain MSB8 (clone 6GC18).

PN MO9744361-A1.

PD 27-NOV-1997.

PF 22-MAY-1997; 97WO-US08793.

PR 22-MAY-1996; 96US-0651572.

PA (RECO-) RECOMBINANT BIOCATALYSIS INC.

PI Lam DE; Mathur EJ;

DR WPI: 1998-018435/02.

DR N-PSDB: AAT94212.

PT Endoglucanase(s), preferably from archaeal bacterium, AEPIT 1a -

PT beta-1,4-glycosidic bonds in cellulose

PS Claim 1; Fig 1T; 16app; English.

CC This protein comprises an endoglucanase of Thermotoga maritima MSB8

CC (clone 6GC18) that is capable of degrading carboxymethylcellulose

CC and of hydrolysing the beta-1,4-glycosidic bonds in cellulose. It

CC has homology to an endoglucanase of archaeobacterium AEPIT1a (see

CC AAW34985). It can be produced from native cells or from recombinant

CC host cells, especially prokaryotic host cells transformed with a

CC plasmid or virus-derived vector including the endoglucanase DNA

CC (see AAT94212). 24 endoglucanases (see AAW34986-W35008) are claimed.

CC They can be used to degrade cellulose for the conversion of plant

CC biomass into fuels and chemicals, for use in detergents, textiles,

CC animal feed, waste treatment, and in the fruit juice and brewing

CC industries for the clarification and extraction of juices.

XX Sequence 778 AA;

Query Match 17.0%; Score 747; DB 19; Length 778;

Best Local Similarity 26.1%; Pred. No. 1.4e-57;

Matches 236; Conservative 149; Mismatches 285; Indels 234; Gaps 34;

QY 3 DIDVAIILKLLTAKVLLACIDFW-----HTKLLPKHGVSLRTPDGN 48

Db 12 EYRVNDDLISRMILEERKVALQSV--WGVELIDERKESREKAKELKNGIQITRRPGST 69

QY 49 GV--RGTKFEN-----GVPA-----AC-----FPGTSLGSTFNOQL 78

Db 70 NLEPDAELVEIQRFLVEETRLGIPAMHIECLITGYMGLGTNFPQALIMASTWBDL 129

QY 79 IBEAGMKMKKALASAHVILGPTTINMORSPLGSGFSGIEDPLLAGAALIRGQ- 137

Db 130 IEKMTTAVEDRKIGAHQGLAPVLDVARDPKRGTEETEFESPYLVARMGVSYKGLQG 189

QY 138 ---STGVQATINHP--CNDQEDRRMAYOSIYTERLARLETALPFOIARVDSOPAFMTAY 193

Db 190 EDIKRGVATVTRHFFAGYSASBEGKMMAPNTNIFEREFEVLEFPFLAAVKEANVLSVMXY 249

QY 194 NGINGVSCSENPRLDGLMKEMGMDGLMSDMYG-----YSTTAVVAG 239

Db 250 SELDGVPCANRKLTLDIRKDMGFEGLVSDYFAVKVLEDYHRIARDKSEARLALBAG 309

QY 240 LDLEMPGPPRFGETLKENVSNKGPPIHVTIDQARAVLOFVKKCAASVTENGPEPTVNN 299

Db 310 IDEVLEPTECY--QYLLKDLVERKGIIEALIDEAVTRVLRK--FMLGLEFN+PYVEVEK 363

QY 300 T--PETAALLRWNGNGIYLVNENNVLPUSKKKTLIYAPNAKOATYHGCSAALRAVY 357

Db 364 AKIESHRDIALEIARKSITLLKND-GILPQNKKALIGPNAGE----- 407

QY 358 AVTPFDGLSKOLETPPSYTVGAYTVTPPIIGQCITPDGAPGRMRVNEPPTPNRQHI 417

Db 408 -----VRNLLGDVMYL-----AAH 421

QY 418 DELFTKTDMHLVDYTHPRADTWYADMEGTATADDCYELGLVYCGTAKAYVDQLVV 477

Db 422 RALLDNIDV---FNPQIPRENYERLKS-----I 449

QY 478 DNATKQVPG--DAFPGSATREETGRINLVKNTKFKIEFGSAPTYTLKGDTIVPGHSL 535

Db 450 EEHMKSIPSVLDAP-----KEEG-----IEFEVAK----- 474

QY 536 RVGCKVYI--DDQAEIKESVALAKEHDQVITC---AGLNADMET--EGADRASKLPLGYLD 589

Db 475 ---GCEVLTGEDRSGFEPAIRIAKSDVAIYVGDKSGLTLDCTTGSRMANKLPGVGE 531

QY 590 QLTADVAANPNVTVMQGTPEEMP-WIDATPAVIOAWYGNENGTNSIADVFGDYNPS 648

Db 532 ELVLEVAKTGRPVVLTITGRPYSLKNVYDKVNAIIVQVLPGEAGRAIVDIIYGVNPS 591

QY 649 GKLSLSPFKRLQDNPAFLNRTGAGRTLYGEDVYVGRYERADKDVNFPFGHGLSYTF 708

Db 592 GKLPISPRAGQIPVHYHYKPGSGSHMHGD-----YDESTKPL-PFGHGLSYTF 644

QY 709 AFSNLVSHKD---GKLSVLSVKNKTSVPGAOVQOLYVKPLQAKINRPVLELGFAP 764

Db 645 EYSNLRIEPEKVPAGVIVKIVDVENIGDRDGEVQVLYT-GREFSVTRPVLELGFAP 703

QY 765 VELQPGETKAVTIEQKVAAYAFDEERD--QWCVEKGDYVIVSDSSAAKGVALLGKFT 823

Db 704 VSLKAKKKTIVRLHMDVLAIV---NRDKLVYVEGEEFVVMGSSS---EDIRIVGSFS 757

QY 824 -VGE 826

Db 758 VVGE 761

RESULT 13

AAW25384

AAW25384 standard; Protein; 744 AA.

ID AAR25384

AC AAR25384;

```

XX 06-JAN-1993 (first entry)
XX Trichoderma reesei B-gal.
XX
XX Beta-galactosidase; filamentous fungi; cellulase; detergent;
XX cellobiosaccharide; grapes; wine; feedstock; biomass; bg11.
XX
XX Trichoderma reesei.
XX
XX W09210581-A.
XX
XX 25-JUN-1992.
XX
XX 10-DEC-1991; 91WO-US09285.
XX
XX 10-DEC-1990; 90US-0625140.
XX
XX (GEMV ) GENENCOR INT INC.
XX
XX Barnett CC, Fowler T, Shoemaker S;
XX
XX WPI: 1992-234636/28.
XX P-PSDB; AAR25384.
XX
XX Extracellular beta-glucosidase expression in filamentous fungi -
XX enhances cellulose degradation in feedstock, biomass and sludge
XX
XX Disclosure; Fig 1; 101pp; English.
XX
XX The beta-galactosidase amino acid sequence was deduced from the DNA
XX sequence of the bg11 gene obtd. using PCR primers based on the N-
XX terminal region of bg11 gene for amplifying a cDNA library to recover
XX a 700 bp bg11 clone which was labelled and used to probe a T.
XX reesei genomic DNA library to recover the full length bg11 gene.
XX Bg11 encodes the 74.3 kD protein beta galactosidase. Transformants
XX of T. reesei can be used to produce fungal cellulase compans. B-gal
XX can be isolated from the culture medium of enriched transformants
XX and added to grapes during wine making, to enhance the potential
XX aroma of the finished wine prod. B-gal may also be used in fruit
XX to enhance the aroma. Enhanced B-gal may be used to degrade
XX cellulosics in feedstock, biomass and sludge. The transformants
XX may be used in detergent compans. to isolate cellobiosaccharides
XX and may be useful as food additives, chemical intermediates, etc.
XX
XX
XX Sequence 744 AA:
XX
XX Query Match 16.7%; Score 731.5; DB 13; Length 744;
XX Best Local Similarity 27.5%; Pred. No. 3.2e-56;
XX Matches 234; Conservative 123; Mismatches 282; Indels 211; Gaps 31;
XX
XX 7 EATLKKTLLEKVDLAGIDFW-----HTKALPKHGVPSLFTGPGNVR---GTFK 55
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 48 KAAIAKLNLDKDGIVSGV-WNGGPCVGTSPASKISYPSLCLDGLGVSTGTAF 106
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 56 FNGVPACFCPCGSLSTFNOTLLEAGKMKREAIKSAHVILGPTIN-MQSPGGRG 114
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 107 TPGVQAA-----STWVNLIRERGQFGEVKAAGIHVLGVAGPLGTPGGGRN 157
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 115 FESIGEDPFLAGAAALIRGIQSTGVQATIKHFLCNDQEDRRMMVQSIIVTERALAEIRA 174
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 158 WEGFGVDPYLTIGMQTIGIQSVQATKHYILNEGLNRETISNPDRITLHELT 217
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 175 LPQIADVRSQPAFMATFATNGINGVSCSENPKYIDMLKREKMGWGLMSDYGITYSTTE 234
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 218 WFPADAVQ-ANAVASYCSTYKNTWACEDQYTLQTLVDLQDGLFPGYVMTDNNAGHTTVO 276
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 235 AVVAGIDLEMPGP-----RFRGETLKFNV-SNGKPEHIVIQRAREVLOFVKKCAASV 288
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 277 SAANSGLDMGMPGDEGNRNKRPALTNVNSNOVPTSVDDM---VTRIIAAVYLTGQ 332
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 289 TENG-PEITVNNTPP--TAALLRKVGNELIVLKNENNVLPISKKKTLIVGPNKQATY 345
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

DB 333 DOAGYPSFNISRNVQGNHKTNRVRAIARDGIVLLKNDANILPLKKPASIADV----- 383
QY 346 HGGGSALRAVYAVTPPDGSLKLETPPSYVATYTPPILEGQ-CLIPDGAFGRRRV 404
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 384 ---GSAAT-----ICNHRNSPSCNDKGC--DDAALGGMW-- 413
QY 405 FNEPPTGNRQIHDELFTTKDMHLDVYYHPKAAADVADMEGTYPADBCITYELGLVVC 464
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 414 ---GSGAVNRPY-----FVAPYAIATRASQCT----- 439
QY 465 GTAKAYVDQQLVVDNATKOYPCDAPFGSATREETGRINLVKNTYKTEFGSAPTYTLK 524
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 440 -----QVLTSLNDNTSSG---ASAARGDAVAIVEITADS----- 470
QY 525 GDTIVPGHSLRVGCKGVDDDAEIEKSVALKKEHQVITCAGLNADWTEGADASMKL 584
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 471 -----GEGTIYEG--NAGDRNND-----PMHNGNA----- 495
QY 585 PGVLQDLIADVAANPNTVVMOTGTP---EEMPMLDATPAVIOAWYGGNETGNSIADV 641
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 496 -----LYQAVGANSNVIVVHSGAILLEQILALPQKAVVMGILPQSGSGLNLDVYL 549
QY 642 FGDYNSGKLSLSPKRLQDNPAFLNR-TEAGRLYGEDVYVGRYEEFADKDVNPPG 700
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 550 WGDVSPSGKLVYTIKAK---SPNDYNTRIYSGSDSFSGLFIDYKHFDDANITPRYEEG 605
QY 701 HGLSYTTFAPFSLSY--SHKDKK-----LSVSLVKNKTSVPGAQVQ 741
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 606 YGLSTYKFPYKSLSTVSTKSGPATGAVVPGPSDLFQNVATIVYDIANSQGVTAQVQ 665
QY 742 LVYK-PLQAKINRPYKELGFAKVELQPGETKAVTIEQEKYVAAYPDEBDQCVK 800
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 666 LYITPSSAPR--TPPKQKRGFAKINLTPGSGTATFNIRRD-LSYMDTASQKVVPSG 722
QY 801 DYEVIYSDSS 810
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 723 SFGISVGAASS 732

```

RESULT 14  
AAB08340  
ID AAB08340 standard; Protein: 744 AA.  
XX  
AC AAB08340;  
XX  
DT 04-DEC-2000 (first entry)  
XX  
DE Amino acid sequence of a beta-glucosidase polypeptide.  
XX  
KW Beta-glucosidase; bg11 gene; filamentous fungus.  
XX  
OS Trichoderma reesei.  
XX  
PN US6103464-A.  
XX  
PD 15-AUG-2000.  
XX  
PF 05-JUN-1995; 95US-0463461.  
XX  
PR 24-MAY-1994; 94US-0248586.  
PR 10-DEC-1991; 91US-0807028.  
PR 10-DEC-1990; 90US-0625140.  
XX  
PA (GEMV ) GENENCOR INT INC.  
XX  
PI Shoemaker S, Barnett CC, Fowler T;  
XX  
DR WPI: 2000-557671/51.  
DR N-PSDB; AAA63953.  
XX  
PT Detecting DNA encoding beta-glucosidase from filamentous fungus  
PT comprises hybridizing fungal DNA with a nucleotide sequence encoding  
PT Trichoderma reesei beta-glucosidase and detecting the hybridized DNA

XX Disclosure: Fig 1; 36pp; English.

CC The present sequence represents a beta-glucosidase. The beta-glucosidase  
CC polypeptide is extracellular or cell wall bound protein encoded by the  
CC bgII gene. The specification describes a method for detecting DNA  
CC encoding beta-glucosidase from a filamentous fungus. The method  
CC comprises contacting the DNA of the fungi with a Trichoderma reesei  
CC bgII polynucleotide, and detecting hybridised DNA sequences. The method  
CC is used for detecting DNA encoding beta-glucosidase from filamentous  
CC fungi. Once homologous genes of other filamentous fungi are cloned,  
CC they can be used to transform the filamentous fungi to overproduce  
CC beta-glucosidase. Alternatively, the cloned bgII genes can be used to  
CC delete or disrupt the bgII genes.

XX Sequence 744 AA:

Query Match 16.7%; Score 731.5; DB 21; Length 744;  
Best Local Similarity 27.5%; Pred. No. 3.2e-56;  
Matches 234; Conservative 123; Mismatches 282; Indels 211; Gaps 31;

```

QY 7 EALIKKLTIAEKVLLAGIDFW-----HTKALPKHGVPSLFTDGPNGVR---GTRF 55
DB 48 KALAKLNIQDKVGIYGVG-WNGGPCVGNTPSPASKISYPSICLDQGPVGVRSTGSTAF 106
QY 56 FNGVPACFPCCGTSLSSTFNQTLLEKAGKMKKEALAKSAHYLGTIN-MORSPILGGRG 114
DB 107 TPQVOAA-----STWDVNLIRRGQFIEGVKASGIHVLIGVAGPLGKTPGGGRN 157
QY 115 FESIGEDPFLAGIAALRGISTGVOATIKFPLNDQEDRBMVQSYTERALREIYA 174
DB 158 WEGFGVDPIYLTGAMQOTINGISVQATAKHYILNEQPLNEFTISSNDDDTLHEIYT 217
QY 175 LPRQIAVRDSQPGAFMTAYNGINGVSCSNPKYIDGMLKREKMGDLINSQWYGTYSIE 234
DB 218 WPFADAVQ-ANVASVMCSYKVNVTWACEDQYTLQTLKQDLGPGVYMTDMAOHTTYQ 276
QY 235 AVAGIDLEMPRP-----RFGETLKFNV-SNGKPIHYIDQARAREVLOFVKKCAASV 288
DB 277 SANGSLMSPEGDENGNNKMLGPAALNAVNSQVPTSRVDM---VRIILAWIITLQ 332
QY 289 TENG-PEYVNNTPPE--TAALLRKVNGEIVLLKNNNNVLPSSKKKTLIVGNARQATY 345
DB 333 DQAGYFPFNISRNQGHKTNVRAIARDGIVLLKNDANIILPKKPASIAVV----- 383
QY 346 HGGGSAALRAYAVTPPDGSLSKOLETPPTVGAATYTPPILEGQ-CLTPDGAQGMKRW 404
DB 384 ---GSAAI-----IGNHARNSPSCNDKGC--DDGALGMGW-- 413
QY 405 FNEPPGPNROHIDELFFETKIDMLVDYVHPKADTWYADMEGYTDADEDCYELGLVVC 464
DB 414 ---GSGAVNPPY-----FVAPYDAINRASSQGI----- 439
QY 465 GTARAYVDOLVDNATKOVPGDAFEGSAPRETGRINLVKNTYKFIKFGSAPYTLK 524
DB 440 -----QVTLNNTNTSSG---ASAARGKVAIVFTADS----- 470
QY 525 GDTIVPBGHSIRVGGCKVYIDQAEIKESVALAKHDVITICAGINADMETEGADRASKML 584
DB 471 -----BEGYITVEG--NAGDRNLD-----PWANGMA----- 495
QY 585 PGVLDLIADVAANPTVVVMOTGTP---EEMFMDATPAVIOAGNGETGNSIADV 641
DB 496 -----LVQAVAGANSNVIVVHVGAILLEQILALPOVAVAVWAGLPQSGESGALVDL 549
QY 642 FGDYVPSGKLSLPPKRLQDNPAFLNFR-TEAGRTLYGEDVYVYRIRYERADKDVNPPFG 700
DB 550 WGDVSPSSKLVYITAK-----SPDNYNTRIVSGSDSEGLEFDYKHFHDANITPREFG 605
QY 701 HGLSTTFAPFNSLVY--SHKDGK-----LSVSLVSKNTGSGVPGAQVQ 741
DB 606 YGLSTTKFNYSRLSVLSTAKSGPATGAVVGGPSDIFQNAVITVDIANSQVYGALEAQ 665

```

```

QY 742 LYVK-PIQAKINRPNKELKGFARVELQPGETKAVTTEDEKRYAAYFDELRQCVKEG 800
DB 666 LYITPSSAPR--TPPKLRGFAKLNLPGSGATFNRRLD-SWWDTSQKAVVPSG 722
QY 801 DYEIVSDSS 810
DB 723 SFGISVGASS 732

```

# RESULT 15

AAV56548  
AAV56548 standard; Protein; 744 AA.

AAV56548;

21-FEB-2000 (first entry)

Trichoderma reesei beta-glucosidase.

Trichoderma reesei; beta-glucosidase; bgII; filamentous fungus;

cellulase; food; flavour; aroma; wine; fruit.

Trichoderma reesei.

US597913-A.

07-DEC-1999.

05-JUN-1995; 95US-0462080.

24-MAY-1994; 94US-0248586.

10-DEC-1991; 91US-0807028.

10-DEC-1990; 90US-0625140.

(GENEV) GENENCOR INT INC.

Shoemaker S, Barnett CC, Fowler T;

WPI: 2000-052536/04.

N-PSDB; AA38879.

Enhancing the flavor of foods using an enzyme overexpressed by a

recombinant filamentous fungi -

Claim 4; Fig 1; 45pp; English.

A method has been developed of enhancing the flavour of foods using  
beta-glucosidase overexpressed by a recombinant filamentous fungi  
selected from the genus *Trichoderma*, *Aspergillus*, *Neurospora*, *Humicola*  
and *Penicillium*. The method comprises: (a) expressing a fungal DNA  
sequence encoding beta-glucosidase in a transformed host using an  
expression vector, where the DNA sequence is capable of being amplified  
by polymerase chain reaction (PCR) with primers GCGTGCCTCCGCGCNG  
(AA38880) and GTTGTTCCTTAACTCTGT (AA38881) under amplification  
conditions of: denaturation at 95 plus degrees Celsius for 10 minutes,  
annealing at 50 plus degrees Celsius for 2 minutes and extension at  
65 plus degrees Celsius for 10 minutes, for 30 cycles; (b) culturing  
the transformants under conditions to permit growth; (c) isolating  
beta-glucosidase produced from the transformants; and (d) adding the  
beta-glucosidase to foods. The method is useful for enhancing the  
flavour or aroma of wine or fruit. The present sequence represents  
beta-glucosidase isolated from *Trichoderma reesei*.

Sequence 744 AA:

Query Match 16.7%; Score 731.5; DB 21; Length 744;  
Best Local Similarity 27.5%; Pred. No. 3.2e-56;  
Matches 234; Conservative 123; Mismatches 282; Indels 211; Gaps 31;

```

QY 7 EALIKKLTIAEKVLLAGIDFW-----HTKALPKHGVPSLFTDGPNGVR---GTRF 55
DB 48 KALAKLNIQDKVGIYGVG-WNGGPCVGNTPSPASKISYPSICLDQGPVGVRSTGSTAF 106

```

QY 56 FNGVPAACEPGCTSLGCTFNMOTLLEAGKMMKEALIAKSAHVILGPTIN-MQSRPLGGRG 114  
Db 107 TFGVQAA-----STWDVNLIREGOFIEEVKASGIIHVILGPAVGLKTPQGGRN 157  
QY 115 FESIGEDPFLAGLGAALIRGISTGVQATIKHFLCNDGEDRRMVOQSTVTERALREIYA 174  
Db 158 WEGFVDPFLTGIAWGQTNGISGVQATAKHYILNEOELNRETISSNPDDRLLHLEYT 217  
QY 175 LPEQIAVRDSQPGAMTAVNGINGVSCSENPXYLDGMLRKEMWDGLIMSDWGYSTTE 234  
Db 218 WEFADAVQ-ANVASVMCSNKNVTWACEDQYTLQTLVAKDQGLFGGYMTDMNAQHTVQ 276  
QY 235 AVVAGLDLEMGPP-----RRGETLKENV-SNGKPIHVIDQRAEVLQFVKCAASGV 288  
Db 277 SANSGLDMSMPTDENGNNRMLGPNALTNVANSNOVPTSRVDM---VTRILAAWYLLGQ 332  
QY 289 TENG-PETTVNTPPE--TAALLRKVNGEIVLKNENNVLPLSKKKTLIGPNAKQATY 345  
Db 333 DQAGYPSFNISANNVGNKTNVRAIARDGIVLLKNDANILPLKRPASTAVV----- 383  
QY 346 HGGGSAALRAYAVVFPDGLSKOLETPPSYTVGAVTVPPILGEO-CLTPDAPGRWRY 404  
Db 384 ---GSAAI-----IGNHARNSPSCNDKGC--DDGALGMGW-- 413  
QY 405 FNEPPTGNRQHIDELFTKTKIDMLVDYIHPKADPTWADMEGTYTADECTYELGLVC 464  
Db 414 ---GSGAVNVPY-----FVAPYDAINFRASSQGT----- 439  
QY 465 GTAKAYVDQVLVDNATKQVPGDAFPGSATREETGRINLVKNNTYKFKIERGSAPTYTLK 524  
Db 440 -----QYTLSTNDTSSG---ASAARGKDAIVFITADS----- 470  
QY 525 GDTIVPGHSLRVGGCKVIDDQAEIEXSVALAKEHDOVILICAGLNADMETEGADRASMKL 584  
Db 471 -----GEGYITVEG--NAGDRNNID-----PMHNGNA----- 495  
QY 585 PGVLDQLADVAANPNVYVVMOTGP---EEMPLDAPPAVIOAMYGNETGNSIADVV 641  
Db 496 -----LVQAAVAGANSNIVVHVSVAIIIEQIILALPOVKAVVWAGLPSQESGNALVDVL 549  
QY 642 FGDNPSGKLSLSPFKRLQDNPAFLNFR-TEAGRTLGEDEVYVGRYEFADKXDNVNPFG 700  
Db 550 WGDVSPSGKLVYITAK-----SPNDYNTRIYVSGSDSEGLFIDYKHFDDANITPRYEPG 605  
QY 701 HGLSYTTFASFNSLV--SHKDGK-----LSVSLSVKNTGSPVGAQVQ 741  
Db 606 YGLSYTKFNYSRLSLSTAKSGPATGAIVVPGPSDLFQNVATVTVDIANSGOVTGAIEVQ 665  
QY 742 LYVK-PLQAKINRPVKELGFAVELQPGETKAVTIEQEKYAAAFDEBRDQWCEYEG 800  
Db 666 LYITYPSSAPR--TPPKQLRGFAKLNTLPGSGTATFNIRRDLL-SYWDYASQKVVVPSG 722  
QY 801 DYEVIYVSDS 810  
Db 723 SFGISVGCASS 732

Search completed: April 26, 2003, 13:08:21  
Job time : 88 secs